SRI International

ON THE RELATION BETWEEN DEFAULT AND AUTOEPISTEMIC LOGIC

Technical Note 407

August 1987
Revised November 1987

By: Kurt G. Konolige, Computer Scientist
Artificial Intelligence Center
Center for Study of Language and Information
Computer and Information Sciences Division

APPROVED FOR PUBLIC RELEASE:
DISTRIBUTION UNLIMITED

This research was supported by the Office of Naval Research under Contract No. N00014-85-C-0251, by subcontract from Stanford University under the Defense Advanced Research Projects Administration, Contract No. N00039-84-C-0211, and by a gift from the System Development Foundation.

SRI International®

333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
NOV 1987		**00-11-1987 to 00-11-1987**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
On the Relation Between Default and Autoepistemic Logic	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **SRI International,333 Ravenswood Avenue,Menlo Park,CA,94025**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**		**46**	

Abstract

Default logic is a formal means of reasoning about defaults: what *normally* is the case, in the absence of contradicting information. Autoepistemic logic, on the other hand, is meant to describe the consequences of reasoning about ignorance: what must be true if a certain fact is *not* known. Although the motivation and formal character of these two systems are different, a closer analysis shows that they share a common trait, which is the indexical nature of certain elements in the theory. In this paper we compare the expressive power of the two systems. First, we give an effective translation of default logic into autoepistemic logic; default theories can thus be embedded into autoepistemic logic. We also present a more suprising result: the reverse translation is also possible, so that every set of sentences in autoepistemic logic can be effectively rewritten as a default theory. The formal equivalence of these two differing systems is thus established. This analysis gives an interpretive semantics to default logic, and yields insight into the nature of defaults in autoepistemic reasoning.

Contents

1 Introduction

Default reasoning can be informally described as the process of jumping to conclusions based on what is normally the case. To say that "power corrupts," for example, is to say that for typical x, in typical situations, x will be corrupted by the exercise of authority.

Default logic [14] is a formalization of default reasoning. An agent's knowledge base (KB), its collection of facts about the world, is taken to be a first-order theory. Default reasoning is expressed by *default rules* of the form

$$\frac{\alpha : M\beta}{\omega}, \tag{1}$$

which can be read roughly as, "If α is provable from the KB, and β is consistent with it, then assume ω as a default." Unlike ordinary first-order inference rules, default rules are *defeasible*: given a KB containing just α, for example, the rule above would allow the inference of ω, but if $\neg\beta$ is added to the KB, then the default rule is no longer applicable. Default rules are thus nonmonotonic inference rules.

In default logic, the default rules operate at a metatheoretic level, as they are not expressed in the language of the KB, and are not inference rules within the KB. Rather, they can be thought of as a means for taking a KB and transforming it into another by the addition of sentences that are not logically derivable from the original. The transformation is defined in terms of a fix-point operator.

This formulation of default reasoning leads us to ask several questions which do not have readily apparent answers. The first concerns the expressiveness of the logic. Certain simple types of defaults can be readily stated; for example, "power corrupts" could be expressed as

$$\frac{Powerful(x) : M\,Corrupt(x)}{Corrupt(x)}, \tag{2}$$

but it is not clear that more complicated constructs could be accommodated. One example is conditional defaults, where a default rule is the conclusion of an implication; another is a default whose consequent is itself a default. Because the default rules are not part of the logical language, there is no obvious, straightforward expression of these concepts.

The second question, related to the first, concerns the semantics of default logic. Because defaults are expressed as inference rules operating in conjunction with a fixed-point construction, the *meaning* of such objects as $M\beta$ is not clear. In some recent work, there have been proposals for a semantics for a restricted class of default theories [8] and for default logic in general [3]. In both cases, the "semantics"

is a reformulation of the KB-transformation induced by the defaults in terms of restrictions on the models of the KB. Although such a reformulation can provide an alternative view of the construction of default extensions, it does not provide a semantics in the sense of providing an interpretation for default rules in a model structure (an *interpretive semantics*). Indeed, because defaults are expressed as inference rules, they are not amenable to interpretation in this fashion.

Our idea in this paper is to define default reasoning within the theory of the KB itself, rather than as a transformation of the KB. If we take the sentences of a KB to be the knowledge or beliefs of an agent, then defaults can be expressed by referring to what an agent *doesn't know*. The default that "power corrupts" could be stated informally as

$$\begin{array}{l}\text{If } x \text{ is powerful, then assume } x \text{ is corrupt} \\ \textit{if nothing known contradicts it.}\end{array} \tag{3}$$

It is easy to see that such reasoning is defeasible in the presence of additional information about the integrity of x. From a formal point of view, it is clear that to assert this statement, the language of the KB must be augmented by a construction that refers to the KB as a whole.

Let us call a theory containing an operator that refers to the theory itself an *indexical theory*. We will use the expression $L\phi$ within a theory to mean that the sentence ϕ is part of the theory. Now we can rephrase the default rule (1) in the following manner, using the operator L:

$$L\alpha \wedge \neg L\neg\beta \supset \omega\,. \tag{4}$$

The intent of a rule of this form is something like: "If α is in the KB, and $\neg\beta$ is *not* in the KB, then ω is true." The negation sign in $\neg\beta$ arises from the use of the provability operator L, the dual of the consistency operator M. Because L is an operator of the KB language, we have been able to express the default within the language of the KB itself, rather than as a metatheoretic construct.

The introduction of an indexical operator is an added complexity, for now we allow our initial KB to contain statements not only about the world, but also about its own contents. Indeed, even interpreting the modal operators of (4) is a problem. Fortunately, the mathematical properties of indexical theories have recently been studied by Moore [13] as a formalization for a another type of nonmonotonic reasoning, called *autoepistemic reasoning*, in which an agent reasons about the relationship of her knowledge to the world. Moore has derived an elegant and natural interpretive semantics for indexical theories incorporating the self-referential operator L. This semantics gives an interpretation to the operator L based on model structures.

We are naturally led to ask what relationship exists between default theories and their corresponding expression in AE logic. Are they essentially different, in the sense that agents using each one would have widely differing sets of beliefs? The answer, which is the main result of this paper, is *no*: default logic and AE logic sanction the same inferences on corresponding initial inputs. This fact has several important consequences. Since default rules are expressible in AE logic, both default and autoepistemic reasoning can be combined within this single formalism. Also, the formal expression of defaults gains the benefits of an interpretive semantics.

A second and more surprising consequence is that AE logic is no more expressive than default logic, even though the L operator is part of the language: there exists a translation from every set of AE logic premises into a corresponding default theory. As we shall see, it is possible by translating the appropriate AE logic statements to construct default theories with the effect of conditional defaults, defaults whose conclusion is a default, and so on. The expression of these concepts is still much more natural in terms of the L operator, but the mathematical properties of the corresponding default theories are the same.

Of independent interest are some results in the theory of AE logic, especially the characterization and equivalence of moderately-grounded and minimal extensions, the definition of strongly-grounded extensions and introspective idempotency, and the relationship of AE logic to the modal logic of weak *S5*.

2 Autoepistemic Logic

Autoepistemic (AE) logic was defined by Moore [13] as a formal account of an agent reasoning about her own beliefs. The agent's beliefs are assumed to be a set of sentences in some logical language augmented by a modal operator L. The intended meaning of $L\phi$ is that ϕ is one of the the agent's beliefs; thus the agent could have beliefs about her own beliefs. For example, consider a space shuttle flight director who believes that it is safe to launch not because of any positive information, but by reasoning that if something were wrong, she would know about it from her engineers. This belief can be expressed using sentences of the augmented language. If P stands for "It is safe to launch the shuttle," then

$$\neg L \neg P \supset P \tag{5}$$

expresses the flight director's self-knowledge. Equation (5) is a logical constraint between a belief state ($L\neg P$) and a condition on the world (P).

The primary focus of AE logic is a normative one: given an initial (or *base*) set of beliefs A about the world, what final set T should an ideal *introspective* agent

settle on? If we restrict ourselves for the moment to languages without the self-belief operator, then clearly an ideal agent should believe all of the logical consequences of her base beliefs, a condition sometimes referred to as *logical omniscience* [5]. More formally, let the expression $\Gamma \models \phi$ mean that the sentence ϕ is logically implied by the set of sentences Γ. Then, if the base set is A, the belief set T of an ideal agent is given by:

$$T = \{\phi \mid A \models \phi\} . \tag{6}$$

The presence of a self-belief operator complicates matters. Because the intended meaning of $L\phi$ depends on the belief set of the agent, the definition of the belief set itself becomes circular, which necessitates the use of a fixed-point equation to define T. In this section we will present this definition and give several alternative formulations that will prove useful.

2.1 Logical preliminaries

We begin with a language $\mathcal{L}$ for expressing self-belief, and introduce valuations of $\mathcal{L}$. The treatment generally follows and extends Moore [13], but differs in two ways. First, the base language is first-order rather than propositional; but this is a minor change, because no quantifying into a modal context is permitted. Second, ideal belief sets are defined with a fixed-point equation over valuations of the language. This definition is equivalent to Moore's original one, but leads to different insights on the nature of the ideal belief set, simpler proofs of many results, and several natural extensions.

Let $\mathcal{L}_0$ be a first-order language with functional terms and a distinguished sentence $\bot$ which is always false (the sentence $\top$ is defined as $\neg\bot$). The normal formation rules for formulas of first-order languages hold. A *sentence* of $\mathcal{L}_0$ is a formula with no free variables; an *atom* is a sentence of the form $P(t_1, \cdots, t_n)$. We extend $\mathcal{L}_0$ by adding a unary modal operator L; the extended language is called $\mathcal{L}$. $\mathcal{L}$ can be defined recursively as containing all the formation rules of $\mathcal{L}_0$, plus the following:

$$\text{If } \phi \text{ is a sentence of } \mathcal{L}, \text{ then so is } L\phi. \tag{7}$$

An expression $L\phi$ is a *modal atom*. Sentences and atoms of $\mathcal{L}_0$ are called *ordinary*. Note that nestings such as $LL\phi$ are modal atoms (and hence sentences) of $\mathcal{L}$. Because the argument of a modal operator never contains free variables, there is no quantifying into the scope of a modal atom, e.g., $\exists x LPx$ is not allowed. A sentence has a *modal depth* n if its modal operators are nested to a depth of n; e.g., $L(P \vee LP)$ has a modal depth of 2. We use the abbreviation $\mathcal{L}_n$ for the set of

all sentences of modal depth n or less. Often we will use a subscript to indicate a subset of sentences based on modal depth; e.g., $\Gamma_n = \Gamma \cap \mathcal{L}_n$.

From the point of view of first-order valuations, the modal atoms $L\phi$ are simply nilary predicates. Our intended interpretation of these atoms is that ϕ is an element of the belief set of the agent. So we will consider valuations of $\mathcal{L}$ to be standard first-order valuations, with the addition of a belief set Γ. The atoms $L\phi$ are interpreted as true or false depending on whether ϕ is in Γ. To distinguish these valuations, we will sometimes call them *L valuations*.

The interaction of the interpretation of L with first-order valuations is often a delicate matter in this paper, and so a perspicuous terminology for talking about L valuations is necessary. In particular, it is often useful to decouple the interpretation of modal and ordinary atoms. First-order valuations are built upon the truthvalues of atoms: for ordinary atoms, truthvalues are given by a structure $\langle U, \varphi, \mathcal{R} \rangle$, where φ is a mapping from terms to elements of the universe U, and $\mathcal{R}$ is a set of relations over U, one for each predicate. We will refer to any such structure as an *ordinary index*, and denote it with the symbol I. Modal atoms are given a truthvalue by a belief set Γ, which is called a *modal index*.

The truthvalue of any sentence in $\mathcal{L}$ can be determined by the normal rules for first-order valuations, given an ordinary and modal index. We write $\models_{I,\Gamma} \phi$ if a valuation $\langle I, \Gamma \rangle$ satisfies ϕ. The valuation rule for modal atoms can be written as

$$\models_{I,\Gamma} L\phi \quad \text{if and only if} \quad \phi \in \Gamma \,. \tag{8}$$

A valuation that makes a every member of a set of sentences true is called a *model* of the set. A sentence that is true in every member of a class of valuations is called *valid* with respect to the class. The following classes of valuations are useful:

$$\begin{array}{rl} \models_\Gamma & \text{valuations with modal index } \Gamma \\ \models & \text{all valuations} \\ \Sigma \models & \text{models of } \Sigma \,. \end{array} \tag{9}$$

A sentence ϕ is a *first-order consequence* (FOC) of a set of sentences Σ if it is true in all models of Σ. Σ is *closed under first-order consequence* if it contains all sentences that are true in all of its first-order models.

2.2 Autoepistemic extensions

Now we return to the original question of what an ideal introspective agent should believe. Obviously, we want to use equation (6), with an appropriate choice for logical implication. Given that the intended meaning of L is *self*-belief, it becomes

obvious that we should consider all models in which the interpretation of $L\phi$ is the belief set of the agent itself; that is, the valuations we consider all have a modal index that is the belief set of the agent. Following Moore, we call such valuations *autoepistemic* (or *AE*), and define the extension of a base set A of beliefs as follows:

DEFINITION 2.1 *Any set of sentences T which satisfies the equation*

$$T = \{\phi \mid A \models_T \phi\}$$

is an autoepistemic extension *of A.*

This is a fixed-point equation for a belief set T, and is a candidate for the belief set of an ideal introspective agent with premises A. It is similar to the belief set definition for a nonintrospective agent (Equation 6) in that it contains A and is closed under first-order consequence. As we will see in later sections, there are other conditions that we may want extensions to satisfy if they are to be considered as ideal belief sets.

Defining AE extensions in this manner gives us an alternative, compact expression of the *stable expansions* of Moore's original exposition [13]. He defines a set of sentences T of $\mathcal{L}$ as *sound* with respect to the premises A if every AE valuation of T (that is, every L valuation with modal index T) that is a model of A is also a model of T. T is *semantically complete* if T contains every sentence that is true in every AE model of T. If T is sound and complete with respect to A, it is called a stable expansion of A.

It is easy to verify from the fixed-point equation that every AE extension is sound and complete with respect to A. Further, any stable expansion of A also satisfies the fixed-point equation. Hence stable expansions are exactly AE extensions.

EXAMPLE 2.1 A base set A may give rise to one or several AE extensions, or to none. As we show below, any set of ordinary sentences has exactly one extension. The extension for the base set $A = \{P\}$ contains all the first-order consequences of P, but no other ordinary formulas. It contains modal atoms of the form $L\phi$, where ϕ is a FOC of A, and $\neg L\psi$, where ψ is not a FOC of A.

Any set of first-order inconsistent sentences gives rise to the same AE extension, containing all sentences. This is the only AE extension that contains the false sentence $\bot$.

The base set $A = \{LP\}$ has no extensions. For suppose T is such an extension; either $P \in T$ or $P \notin T$. Clearly the latter cannot be the case, for then for *any* sentence ϕ, $A \models_T \phi$ (because $\models_{I,T} A$ is false for any I). Now suppose

$P \in T$. In this case, we can construct an interpretation that satisfies A but falsifies P, namely one in which I makes P false. Therefore it cannot be that $A \models_T P$, and so P is not in T, a contradiction.
The base set $\{LP \supset P\}$ has two extensions, one of which contains P, and the other of which does not.
The base set $\{\neg LP \supset Q,\ \neg LQ \supset P\}$ has two extensions; in one of them, LP is true and LQ is not, and in the other the reverse.

Suppose an agent has only ordinary sentences in her base set A. These sentences determine a unique extension for the agent. This proposition was proven independently by Marek [9]. (To make the development of this paper clearer, we include proofs of propositions in an appendix).

PROPOSITION 2.1 *If A is a set of ordinary sentences, it has exactly one AE extension T. T_0 is the first-order closure of A.*

We now consider an alternative semantic characterization of AE extensions. In Definition 2.1, an extension T is defined using the operator $\models_T$, which incorporates T itself. However, if the definition used the simple validity operator $\models$, self-reference would be eliminated, and a proof-theoretic analog for $\models$ could be used. Note that in $\models_T$, the modal index of the interpretation gives truthvalues to all atoms of the form $L\phi$, according to whether ϕ is in T or not. Let LT stand for the set of formulas $\{L\phi \mid \phi \in T\}$, and $\neg L\overline{T}$ for $\{\neg L\phi \mid \phi \notin T\}$. The sets LT and $\neg L\overline{T}$ are satisfied only by the interpretations whose modal index is T; hence $LT \cup \neg L\overline{T} \models \phi$ if and only if $\models_T \phi$. This gives the following alternative characterization of extensions:

PROPOSITION 2.2 *A set T is an AE extension of A if and only if it satisfies the equation*

$$T = \{\phi \mid A \cup LT \cup \neg L\overline{T} \models \phi\}\ .$$

Essentially, the self-referential part of the definition has been transferred from the implication operator to the set of assumptions LT and $\neg L\overline{T}$. Note that there is a trade-off between the strength of assumptions and of the implication operator: in substituting the weak operator $\models$ for $\models_T$, we were forced to introduce the strong assumptions LT and $\neg L\overline{T}$ to eliminate unwanted models.

Moore has called belief sets defined by the above equation *grounded in A*, because they are derived from A and assumptions about self-belief. This notion of groundedness is a fairly weak one, in that the allowable assumptions are very liberal — an agent is able to assume a self-belief in any proposition as a basis for the

derivation of beliefs. In later sections we will define two strengthenings of groundedness; to distinguish Moore's definition, we will call extensions *weakly grounded* if they obey the equation of Proposition 2.2.

It is possible to modify the above equation by strengthening the implication operator (still avoiding self-reference) and weakening the assumptions. This gives us yet another semantic characterization of AE extensions. However, we first need to introduce and analyze a special type of belief set, called a *stable set.*

2.3 Stable sets

Following Stalnaker [16], we call a belief set Γ *stable* if it satisfies the following three properties:

1. Γ is closed under first-order consequence.[1]
2. If $\phi \in \Gamma$, then $L\phi \in \Gamma$.
3. If $\phi \notin \Gamma$, then $\neg L\phi \in \Gamma$.

Given Proposition 2.2, it is clear that AE extensions must be stable sets. They are closed under first-order consequence because they are defined using the operator $\models$; and the presence of LT and $\neg L\overline{T}$ guarantees that properties (2) and (3) above are satisfied. Thus we have:

PROPOSITION 2.3 (MOORE) *Every AE extension of A is a stable set containing A.*

The strict converse of this proposition is not true, since there can be stable sets containing A that are not AE extensions of A. The simplest example is $A = \{LP\}$, which has no AE extension (see Example 2.1). Yet there are many stable sets that contain LP.

A partial converse is available if we consider stable sets as AE extensions of their own ordinary sentences.

PROPOSITION 2.4 *Every stable set Γ is an AE extension of Γ_0.*

Stable sets are thus AE extensions of their ordinary sentences. From Proposition 2.1, we know that every such AE extension is unique; hence every stable set is uniquely determined by its ordinary sentences.

[1]Stalnaker considered propositional languages and so used tautological consequence.

PROPOSITION 2.5 (MOORE) *If two stable sets agree on ordinary formulas, they are equal.*

The set of ordinary formulas contained in a stable set is closed under first-order consequence. Different stable sets thus have different sets of first-order (FO) closed ordinary formulas. We now show that stable sets cover the sets of FO-closed ordinary formulas; that is, every such FO-closed set is the ordinary part of some stable set.

PROPOSITION 2.6 *Let W be a set of ordinary formulas closed under first-order consequence. There is a unique stable set Γ such that $\Gamma_0 = W$. W is called the* kernel *of the stable set.*

The stable set whose kernel contains the element $\bot$ is the set of all sentences of $\mathcal{L}$. This is the unique inconsistent stable set.

We are now ready to give a third semantic characterization of AE extensions. Since AE extensions are stable, let us consider restricting the range of modal indices on the logical implication operator to just stable sets; we indicate this by $\models_{SS}$. From Proposition 2.5, we know that the ordinary formulas of a stable set uniquely determine it. As usual, let T_0 be the set of ordinary formulas of T, and $\overline{T}_0$ the set of ordinary formulas *not* in T. Then, if T is stable, it must be the case that $\models_T$ is equivalent to $LT_0 \cup \neg L\overline{T}_0 \models_{SS}$, because LT_0 and $\neg L\overline{T}_0$ specify only those models in which the modal index is the unique stable set containing exactly the ordinary formulas T_0. This suggests how we can replace $\models_T$ in the definition of AE extensions.

PROPOSITION 2.7 *A set T is an AE extension of A if and only if it satisfies the equation*

$$T = \{\phi \mid A \cup LT_0 \cup \neg L\overline{T}_0 \models_{SS} \phi\} .$$

By using a stronger type of implication ($\models_{SS}$ over stable sets), we have been able to eliminate all self-referential assumptions except for those involving the ordinary formulas of T. Proposition 2.7 also hints that the nesting of L operators gives no extra expressive power to the language, since stable sets are characterized by giving the sets LT_0 and $\neg L\overline{T}_0$. Indeed this is so, and we will prove it in section 3, when we have introduced proof-theoretic analogs to the semantic fixed-point equations.

2.4 Moderately-grounded extensions

One way of evaluating the fixed-point equations in Propositions 2.2 and 2.7 is by the type of reasoning they sanction for introspective agents. So, according to Proposition (2.2), an agent is justified in believing all of the first-order consequences of her base set A and the assumptions LT and $\neg L\overline{T}$. As we have noted, this is a fairly weak groundedness condition, and we might want the belief sets of ideal reasoning agents to obey stronger constraints. Consider, for example, the base set $A = \{LP \supset P\}$. A has two AE extensions, which we call T and T' (see Example 2.1). T contains P and LP, while T' does not contain P, but has $\neg LP$. The difference between these extensions lies in whether LP is introduced as an assumption in the fixed-point equation of Proposition 2.2. For the belief set T, the agent's belief in P is grounded in her assumption that she believes P. If she chooses to believe P, she is justified in believing it precisely because she made it one of her beliefs. This certainly seems to be an anomolous situation, since the agent can, simply by choosing to assume a belief or not, be justified in either believing or not believing a fact about the real world.

We would like to define a stronger notion of groundedness to eliminate this circularity. Now consider the belief set definition given in Proposition 2.7:

$$T = \{\phi \mid A \cup LT_0 \cup \neg L\overline{T}_0 \models_{SS} \phi\} .$$

The set of ordinary sentences in the belief set is T_0. LT_0 is the assumption that the agent believes all of these sentences. There would be no circular justifications if we replace LT_0 by LA in the fixed-point definition. In this way we are assured that the derivation of facts about the world does not depend on the assumption of belief in those facts. The assumption of A is necessary because an ideally introspective agent should at least believe that her base beliefs are beliefs.

From this discussion, we can define the following notion of *moderately grounded.*

DEFINITION 2.2 *A set of sentences* Γ *is* moderately grounded in A *if it obeys the equation*

$$\Gamma = \{\phi \mid A \cup LA \cup \neg L\overline{\Gamma}_0 \models_{SS} \phi\} .$$

Sets that are moderately grounded in A are also AE extensions of A, as we will shortly show. However, not every AE extension is moderately grounded.

EXAMPLE 2.2 The base set $A = \{LP \supset P\}$ has two extensions, but only one of them is moderately grounded. The extension containing P cannot be moderately grounded, because P cannot be derived without the assumption of LP.

A more complicated case is the base set $A = \{LP \supset Q,\ LQ \supset P\}$. Again there are two extensions, one containing the ordinary formulas P and Q, and one without them. For the former, LP and LQ must be assumed together in order to justify P and Q. Because they cannot be derived without this assumption, this extension is not moderately grounded.
All extensions of the set of ordinary formulas A are moderately grounded, because every $\phi \in T_0$ is in the first-order closure of A and so in the stable set containing LA.

Moderately grounded sets are conservative in what they assume about the world, given the base beliefs. As shown in example 2.2, the base set $\{LP \supset P\}$ has only one moderately grounded extension, for which P is not a belief. In fact, moderate groundedness is closely related to another concept, the *minimality* of ordinary sentences in an extension.

DEFINITION 2.3 *An AE extension T of A is* minimal for A *if there is no stable set S containing A such that $S_0 \subset T_0$.*

Minimal extensions guarantee that there is no other possible belief set that makes fewer assumptions about the world, while at the same time containing A and satisfying the stability conditions for introspection. Minimal extensions are thus appealing candidates for ideal introspective belief sets.

EXAMPLE 2.3

The base set $A = \{LP \supset P\}$ has a single minimal extension, the one that doesn't contain P. Note that there can be more than one minimal extension for a given base set: $A = \{\neg LP \supset Q, \neg LQ \supset P\}$ has two extensions, both of which are minimal for A. But a base set which has extensions does not necessarily have any minimal extensions: the base set $A = \{LP \supset P,\ LP \supset Q,\ LQ\}$ has one extension containing both P and Q, but no minimal extensions, since the stable set whose kernel is the first-order consequences of Q contains A, but is not itself an AE extension of A.

We now prove that, in fact, the minimal AE extensions of A are exactly the sets moderately grounded in A. Thus we have two independent motivations for choosing moderate groundedness as a condition for ideal belief sets.

PROPOSITION 2.8 *A set of sentences is moderately grounded in A if and only if it is a minimal AE extension of A.*

There is an interesting connection between stable sets and minimal AE extensions. By analogy with Definition 2.3, we define a minimal stable set for A as follows:

DEFINITION 2.4 *A stable set S is* minimal for A *if S contains A and there is no other stable set S' containing A such that $S'_0 \subset S_0$.*

From the definitions, it is obvious that every minimal AE extension for A is a minimal stable set for A. The converse of this is not true, since there are minimal stable sets for A which are not AE extensions of A (see the discussion following Proposition 2.3). However, if a minimal stable set for A *is* an AE extension, then it must be a minimal extension, since there are no other stable set containing A with a smaller kernel.

3 Proof Theory

We now examine proof-theoretic analogs to the semantics of AE extensions. The immediate reason for this examination is to establish a normal form for base sets that will be useful in proving the correspondence between AE and default extensions. A longer-term goal, and one which we will not pursue here, is to understand how an agent can reason about and justify her own beliefs, starting from an initial set, by using rules of inference.

3.1 Proof-theoretic fixed-points

The simplest analog is to replace the logical implication operator $\models$ in Proposition 2.2 by a first-order deduction operator $\vdash$. There are sound and complete systems of first-order deduction, that is, the derivations of such systems are exactly the first-order consequences. So we have the following proposition:

PROPOSITION 3.1 (MOORE) *A set T is an AE extension of A if and only if it satisfies the equation*

$$T = \{\phi \mid A \cup LT \cup \neg L\overline{T} \vdash \phi\} .$$

Moore originally proved this theorem for the case of a propositional base language. It gives a completely proof-theoretic characterization of AE extensions. Unfortunately, it is still a fixed-point equation, and so does not yield a constructive method for finding extensions.

Finding a deductive analog to the operator $\models_{SS}$ is more difficult, because it involves logical implication over modal indices that are stable sets. The correct logic is a modal system known as *K45*. For those familiar with modal logic, *K45* is a weakening of the well-known modal system *S5*; it is appropriate for belief, as opposed to knowledge, because beliefs can be false. The development of this result is somewhat long, and involves exploring the connection between stable sets and the possible-worlds semantics for *S5*.

Possible worlds have been used as a semantical basis for a variety of epistemic logics (those dealing with knowledge or belief).[2] In this approach, an agent's beliefs are represented by a set of worlds W, those that are *compatible with* her beliefs. Each world contains a first-order valuation.

Now suppose an agent has a base set A consisting entirely of ordinary sentences. We define the set of worlds W compatible with A as all worlds with valuations that make every element of A true. Because each world has a first-order valuation, all of the first-order consequences of A will also be true at each world. On the other hand, for every sentence ϕ which is not a first-order consequence of A, there must be some world in which $\neg\phi$ is true, or else ϕ would also be a belief.

This takes care of beliefs which are ordinary formulas; what about self-beliefs? Since we assume that the agent is ideally introspective, $L\phi$ should be a belief (and hence true in all worlds) just in case ϕ is a belief. So the semantics of belief atoms is given by:

$$L\phi \text{ is true in } w \quad \text{iff} \quad \forall w' \in W \,.\, \phi \text{ is true in } w' \,. \tag{10}$$

A structure defined by a set of possible worlds and the truth-recursion rule (10) is an *S5 interpretation*. Let us call the set of sentences true at every world in W an *S5 set*, and if W was generated by a set of ordinary sentences A, an *S5 set for* A. These sets have the following properties:

1. If W was generated by A, the ordinary formulas in the set are just the first-order consequences of A.

2. The set is closed under first-order consequence.

3. If ϕ is in the set, then $L\phi$ must also be, according to (10).

4. If ϕ is not in the set, then $\neg L\phi$ must be, because there is some world at which $\neg\phi$ is true.

These are exactly the conditions for stable sets, so every *S5* set must be stable. The converse is also true, namely, every stable set is an *S5* set. Let A be the kernel

[2] A good overview for computer science applications is given by Halpern and Moses [4].

of a stable set S. There is an $S5$ set S' for A, which by the above conditions is also a stable set whose kernel is A. From Proposition 2.5, S and S' are identical. Hence we have the following equivalence between stable sets and $S5$ sets:

PROPOSITION 3.2 *A set is stable if and only if it is an S5 set.*

The relationship between stable sets and $S5$ structures has been noted by others, and a version of this proposition was originally proven by Moore, Halpern and Moses, and Fitting (see [11, p. 7]).

The equivalence between $S5$ sets and stable sets leads to an alternative form of the $\models_{SS}$ operator. Let us define an $S5^+$ *valuation* $\langle w, W\rangle$, where w is a possible world, and W is a set of possible worlds (perhaps containing w), by the truth-recursion rules:

$$\begin{array}{lll} \models_{\langle w,W\rangle} \phi & \text{iff} & \phi \text{ is true in } w, \text{ for ordinary } \phi \\ \models_{\langle w,W\rangle} L\phi & \text{iff} & \text{for all } w' \in W,\ \models_{\langle w',W\rangle} \phi \end{array} \tag{11}$$

If a sentence ϕ is valid in all $S5^+$ valuations, we write $\models_{S5^+} \phi$.[3]

PROPOSITION 3.3 *For any $\phi \in \mathcal{L}$, $\models_{S5^+} \phi$ if and only if $\models_{SS} \phi$.*

We now have equivalence between validity in possible-worlds structures, and validity in stable-set L valuations. There is a large body of literature about possible-worlds structures which we can draw on to derive the correct proof theory of $S5^+$ valuations. First, we must relate $S5^+$ valuations to the usual form of possible-worlds semantics, Kripke stuctures.

A Kripke structure $\langle w, W, R\rangle$ consists of a distinguished possible world, a set of possible worlds (containing w), and a binary relation R among the elements of W (the *accessibility relation*). The truth-recursion rule is:

$$\begin{array}{lll} \models_{\langle w,W,R\rangle} \phi & \text{iff} & \phi \text{ is true in } w, \text{ for ordinary } \phi \\ \models_{\langle w,W,R\rangle} L\phi & \text{iff} & \text{for all } w' \text{ such that } wRw',\ \models_{\langle w',W,R\rangle} \phi \end{array} \tag{12}$$

Different classes of Kripke structures are generated by restrictions on the form of the accessibility relation. The condition which corresponds to $S5^+$ valuations is that R be transitive and euclidean (aRb and aRc implies bRc); let us call these TE valuations. We now prove that TE and $S5^+$ valuations are the same.

[3] $S5^+$ valuations are closely related to the model descriptions of Levesque [7]. The only significant difference appears to be that he assumes W is always nonempty; this corresponds to eliminating the inconsistent $S5$ set.

LEMMA 3.4 *Let R be an equivalence relation on W, and let the successors of w be the subset $W' \subseteq W$. Then the Kripke valuation $\langle w, W, R\rangle$ is equivalent to the $S5^+$ valuation $\langle w, W'\rangle$.*

PROPOSITION 3.5 *A valuation is an $S5^+$ valuation if and only if it is a TE valuation.*

The correct proof theory for Kripke models whose accessibility relation is transitive and euclidean is a system called *K45*, or *weak S5* (see Chellas [2]). Here is the axiomatization for propositional $\mathcal{L}_0$.

DEFINITION 3.1 *By the system propositional K45 we mean the following set of axioms and inference rules:*

all tautologies	
$L(\phi \supset \psi) \supset (L\phi \supset L\psi)$	*Dist*
$L\phi \supset LL\phi$	4
$\neg L\phi \supset L\neg L\phi$	5
$\dfrac{\phi \quad \phi \supset \psi}{\psi}$	*MP*
$\dfrac{\phi}{L\phi}$	*Nec*

The distribution schema (*Dist*) says that *K45*-theorems are closed under modus ponens. Axiom 4 states that if ϕ is true at every world, then so is $L\phi$; and Axiom 5 that if ϕ isn't true at some world, $\neg L\phi$ must be true at every world. Modus ponens and necessitation are the rules of inference. Necessitation and the distribution schema ensure that all worlds are closed under tautologous consequence.

For a first-order base language, we would modify the provision of all tautologies to a suitable generator of first-order valid sentences. We write $\Gamma \vdash_{K45} \phi$ if $(\gamma_1 \wedge \gamma_2 \wedge \cdots \wedge \gamma_n) \supset \phi$ is a theorem of *K45*, for some finite subset $\{\gamma_1 \ldots \gamma_n\}$ of Γ. It is known that *K45* is sound, complete, and compact with respect to TE valuations; hence $\Gamma \models_{\text{TE}} \phi$ if and only if $\Gamma \vdash_{K45} \phi$.

Now we can give the proof-theoretic analogs to Propositions 2.7 and 2.8.

PROPOSITION 3.6 *A set T is an AE extension of A if and only if it satisfies the equation*

$$T = \{\phi \mid A \cup LT_0 \cup \neg L\overline{T}_0 \vdash_{K45} \phi\}\,.$$

It is a minimal (moderately-grounded) extension of A if and only if it satisfies the equation

$$T = \{\phi \mid A \cup LA \cup \neg L\overline{T}_0 \vdash_{K45} \phi\}\,.$$

3.2 Normal form

Normal form significantly reduces the conceptual complexity of AE sentences, since we need not be concerned with nested modal operators. It is essential to the notion of *strong groundedness* in the next subsection, and to the translation of AE logic into default logic.

The following two facts about *K45* theories are useful in establishing a normal form:

1. Every AE sentence is equivalent to a sentence containing modal atoms only of the form $L\phi$ or $\neg L\phi$, where ϕ is an ordinary sentence.

2. $L\phi \wedge L\psi$ is equivalent to $L(\phi \wedge \psi)$.

The first of these facts enables us to consider only base sets A drawn from $\mathcal{L}_1$. As we hinted in the last section, the nesting of L operators lends no extra expressive power to the language, since they can always be re-expressed in terms of $\mathcal{L}_1$.

In deriving a normal form for a set of sentences A, we first show that every sentence from $\mathcal{L}_1$ has an equivalent form in which no modal operator appears in the scope of a quantifier, i.e., it is a boolean combination of modal atoms and ordinary sentences.

PROPOSITION 3.7 *Every sentence of $\mathcal{L}_1$ is equivalent to a sentence of the form*

$$\begin{array}{ll} (L_1 \vee \omega_1) & \wedge \\ (L_2 \vee \omega_2) & \wedge \\ \cdots & \wedge \\ (L_n \vee \omega_n) & \end{array}$$

where each L_i is a disjunction of modal literals on ordinary sentences, and each ω_i is ordinary.

Next, we show that any modal atom with nested modal operators is equivalent to a sentence from $\mathcal{L}_1$.

PROPOSITION 3.8 *Every modal atom $L\phi$, where ϕ is from $\mathcal{L}_1$, is equivalent to a sentence of $\mathcal{L}_1$.*

EXAMPLE 3.1 Let us reduce the following sentence to one from $\mathcal{L}_1$.

$$P \wedge \neg L(L\neg Q \vee LQ) \supset \neg L\neg Q \wedge \neg LQ$$

Concentrating on the second part of the antecedent, we have:

$$\begin{aligned} \neg L(L\neg Q \vee LQ) &\equiv_{K45} \\ \neg(L\neg Q \vee LQ) &\equiv_{K45} \\ \neg L\neg Q \wedge \neg LQ & \end{aligned}$$

Substituting into the original sentence, we get:

$$P \wedge \neg L\neg Q \wedge \neg LQ \supset \neg L\neg Q \wedge \neg LQ \quad ,$$

which is just a tautology.

We are now ready to define normal form for sets of AE sentences. By successively applying Proposition 3.8 to a sentence of $\mathcal{L}$, all nested modal operators can be eliminated. Using Proposition 3.7 and the fact that $L\phi \wedge L\psi \equiv L(\phi \wedge \psi)$, any sentence of $\mathcal{L}_1$ can be converted into a set of simple disjunctive sentences.

PROPOSITION 3.9 *Every set A of $\mathcal{L}$-sentences has a K45-equivalent set in which each sentence is of the form*

$$\neg L\alpha \vee L\beta_1 \vee \cdots \vee L\beta_n \vee \omega \ , \tag{13}$$

with α, β_i, and ω all being ordinary sentences. Any of the disjuncts, except for ω, may be absent.

3.3 Strong groundedness

Consider the following base set:[4]

$$\begin{aligned} &\neg LP \supset Q \\ &LP \supset P \end{aligned} \tag{14}$$

This base set has *two* moderately grounded extensions, which we call T and T'. T contains Q and $\neg LP$, while T' has P and $\neg LQ$. It is easy to see how T is grounded: assuming $\neg LP$, Q is derivable from the base set. On the other hand, it is not clear how, in T', P can be derived from the base set and $\neg LQ$. But recall that in the definition of moderately grounded (2.2), both the base set A and LA are present. In $K45$, $L(\neg LP \supset Q)$ is equivalent to $\neg LQ \supset LP$; from $\neg LQ$, it is possible to derive LP; and from $LP \supset P$, we arrive at P.

[4]This example was suggested by Michael Gelfond and Halina Przymusinska in a private communication.

There is something curious about this reasoning, because LP is derived before P is. It is as if an agent, in the course of reasoning, arrives first at the point of believing "I believe that P," without first having come to believe P itself. In defining moderately-grounded extensions, we eliminated one possible way that this could occur, namely using the *assumption* of LP to justify belief in P. But there are other means by which an agent might arrive at a belief in LP before P, as the above example shows. Here it is the assumption of $\neg LQ$ that leads to the derivation of LP, without having derived P.

So we seek a further notion of groundedness, which we call *strong groundedness*, in which every ordinary sentence ϕ has a derivation that does not depend on $L\phi$. The most straightforward approach would be to exclude sentences such as $LP \supset P$ above, which clearly expresses the derivation. The problem here is that $LP \supset P$ may not be explicitly represented: we could replace it, for example, by the chain $LP \supset Q_1$, $LQ_1 \supset Q_2$, ..., $LQ_n \supset P$, or even more complicated constructions, so just ruling out sentences of the form $LP \supset P$ will not suffice.

The solution is to break the derivation of ϕ from $L\phi$ by never allowing $L\phi$ to be derived, except as a consequence of the derivation of ϕ. There is a simple way to accomplish this. Note that in the above example, T' does not contain Q. Hence the first sentence of the base set, $\neg LP \supset Q$, is not used to derive Q. Rather, given $\neg LQ$, it can be used (via $L(\neg LP \supset Q) \equiv \neg LQ \supset LP$) to derive LP. Now suppose a base set A is in normal form, so all sentences are of the form $\neg L\alpha \vee L\beta_1 \vee \cdots \vee L\beta_n \vee \omega$. We want this sentence to be used only for the derivation of ω, the ordinary formula. We will call an extension *strongly grounded* if all of its base sentences are used in this way.

DEFINITION 3.2 *Let A be a set of AE sentences in normal form, and let T be an extension of A. Let A' be the set of sentences of A whose ordinary part is contained in T. Then T is* strongly grounded *in A if and only if*

$$T = \{\phi \mid A' \cup LA' \cup \neg L\overline{T}_0 \models_{SS} \phi\} .$$

Strongly grounded extensions have some curious properties. First, we show that every strongly grounded extension is moderately grounded (and hence minimal).

PROPOSITION 3.10 *If T is a strongly grounded extension of A, it is moderately grounded in A.*

Not every moderately-grounded extension is strongly-grounded; the minimal extension T' in the example above is not strongly-grounded.

One of the nice properties of moderate groundedness is that it is insensitive to the syntactic form of the base set. If two base sets A and A' are $K45$-equivalent, then they have the same moderately-grounded extensions. However, they do not necessarily have the same strongly-grounded extensions. For example, consider the following two equivalent normal-form base sets:

$$\begin{aligned} A &= \left\{ \begin{array}{l} P \\ LP \vee \bot \end{array} \right\} \\ A' &= \left\{ \begin{array}{l} \neg LP \vee P \\ LP \vee \bot \end{array} \right\} \end{aligned} \tag{15}$$

They both have one extension, which is moderately grounded ($LP \vee \bot$ is equivalent to LP). This extension is strongly grounded for A, but not for A', since eliminating $LP \vee \bot$ leaves no way to derive P.

So strong groundedness is at least partially a *syntactic* property of AE base sets, depending on the exact form of the sentences. This seems inevitable if we want to formalize the notion that ϕ be derived independently of $L\phi$, since the notion of derivation itself is a syntactic one. This idea — that the *form* of the sentences in a KB is important for the derivation of answers from the KB — is neither unfamiliar nor undesirable, and in fact has been at the center of the earliest uses of specialized deductive mechanisms in AI. There is even a name for systems of this sort: *procedural deductive system* (see Moore [12]). The main characteristic of such systems is the presence of several syntactic varieties of implication, all of which have the same semantics, but different procedural interpretations in derivations: forward chaining, backward chaining, and the like. The form in which an implication is expressed has a great influence on the derivational behavior of the system. In like manner, the exact expression of self-belief in AE sentences can have an effect on whether an extension is strongly grounded or not.

Strongly grounded extensions enjoy one useful property that moderately grounded ones do not: they are insensitive to the addition of the schema $L\phi \supset \phi$ to the base set. The only way $L\phi$ could be derived in a strongly-grounded extension is if ϕ is derived independently; hence $L\phi \supset \phi$ will never be used to derive ϕ, and cannot cause the derivation of any new ordinary sentences. Thus T is a strongly-grounded extension of A if and only if it is a strongly grounded extension of $A \cup L\phi \supset \phi$. We call this property *introspective idempotency.*

3.4 Summary of groundedness results

There are three notions of groundedness for AE extensions. Every AE extension T of a base set A is weakly grounded, that is, it satisfies the equation

$$T = \{\phi \mid A \cup LT \cup \neg L\overline{T} \models \phi\} .$$

Moderately-grounded extensions eliminate all assumptions of positive modal atoms, and use a stronger notion of logical consequence:

$$T = \{\phi \mid A \cup LA \cup \neg L\overline{T}_0 \models_{SS} \phi\} .$$

All moderately-grounded extensions are weakly grounded, but the converse is not necessarily true. An alternate characterization of moderate groundedness is provided by a minimality condition: the moderately-grounded extensions of A are just those extensions which are minimal in their kernels (see Proposition 2.8).

Strong groundedness is partially a syntactic concept, based on the normal form for AE sentences. Let A' be all sentences of A in normal form whose ordinary sentence is contained in T. Then T is strongly grounded if

$$T = \{\phi \mid A' \cup LA' \cup \neg L\overline{T}_0 \models_{SS} \phi\} .$$

All strongly-grounded extensions are moderately grounded, but the converse is not necessarily true.

4 Default Logic

We briefly review the logic of default theories. As defined by Reiter [14], a default theory is a pair $\langle W, D\rangle$, where W is a set of first-order sentences and D is a set of defaults, each of which has the form

$$\frac{\alpha : M\beta_1, M\beta_2, \ldots M\beta_n}{\omega} .$$

A default d is satisfied by a set of sentences Γ if either (1) α is not in Γ or some $\neg\beta_i$ is in Γ (the premisses of the rule are not satisfied), or (2) ω is in Γ (the conclusion is satisfied). A default extension of $\langle W, D\rangle$, informally, is a minimal set of sentences containing W, closed under first-order consequence and satisfying all the defaults D.

If none of α, β_i, or ω contain free variables, then the default is called *closed.* An open default is treated as a schema for the set of closed defaults that are its

substitution instances. We thus need consider only closed defaults, as long as we allow default theories to contain a denumerably infinite set of them.

Default extensions share many of the properties of AE extensions. There may be one or many extensions of a default theory, or none. The following examples are analogous to the AE extensions in Example 2.1.

EXAMPLE 4.1 The default extension for the theory $\langle\{P\},\emptyset\rangle$ (no defaults) is exactly the first-order consequences of P.
The theory $\langle\emptyset, P : /P\rangle$ has one extension, the set of all first-order valid sentences. P is *not* an element of this extension. This differs from the AE base set $\{LP \supset P\}$ which has an extension containing P.
The theory $\langle\emptyset, \{M\neg P/Q,\ M\neg Q/P\}\rangle$ has two extensions; in one of them, P is true and Q is not, and in the other the reverse.

These examples are instructive by comparison to AE extensions. If the theory $\langle W, D\rangle$ contains no defaults (D empty), then there is exactly one extension, which is the first-order part of the AE extension of W. In general, a default of the form $\alpha : M\beta/\omega$ corresponds to the AE sentence $L\alpha \wedge \neg L\neg\beta \supset \omega$; thus, in the third default theory of the example, there are two default extensions, corresponding to the first-order parts of the two AE extensions of $\{\neg LP \supset Q,\ \neg LQ \supset P\}$. However, note the difference in the case of the second default theory of this example. The default $P : /P$ has only one extension, in which P does not appear. The AE set $\{LP \supset P\}$ has *two* extensions; the one in which P appears arises from the ability of AE extensions to support circular justifications (assuming LP, the sentence $LP \supset P$ gives a derivation of P). So although it appears that default extensions have corresponding AE extensions for a suitable transformation of the defaults, not all AE extensions will have corresponding default extensions. In fact, as we show below, default extensions correspond to *strongly grounded* AE extensions.

Default extensions are the fixed points of an operator $\Gamma(V)$. This operator is meant to formalize the informal criteria given above for the extensions of $\langle W, D\rangle$, namely, it should contain W, be closed under first-order consequence, and satisfy all of D. Let V be an arbitrary set of first-order sentences. Then $\Gamma(V)$ is the smallest set satisfying the following properties:

D1. $W \subseteq \Gamma(V)$

D2. $\Gamma(V)$ is closed under first-order consequence.[5]

[5]In the original definition, this is stated in terms of deduction rather than logical consequence.

D3. If $\alpha : M\beta/\omega \in D$, $\alpha \in \Gamma(V)$, and $\neg\beta \notin V$, then $\omega \in \Gamma(V)$.

Extensions are fixed-points of Γ, i.e., any set E satisfying $E = \Gamma(E)$. As a fixed-point definition, it is similar to the fixed-point account of minimal AE extensions (Proposition 2.8). The parameter of $\Gamma(V)$ essentially fills the role of the assumptions $\neg L\overline{T}_0$, since $\neg\beta$ must *not* be present in order for the default to be satisfied. Minimality is part of the definition of $\Gamma(V)$ (the *least set* satisfying the conditions D1–D3); if it were excluded, then default extensions corresponding to nonminimal AE extensions would be present.

5 Default and AE Extensions

In this section we explore the relationship between default and AE extensions.

5.1 Translations

In standard mathematical reasoning, to compare a formal systems Q_1 to another system Q_2 it is necessary to provide a sentence-to-sentence translation from the language of Q_1 to the language of Q_2 (see Boolos [1, p. 46]) The difficulty here is that the *language* of the two formalisms is different: default theories are first-order but contain inference rules with a metatheoretic operator M, while AE logic has a modal operator L in the language itself. How do we go about comparing the two?

Both formalisms have in common the notion of an *extension*. In AE logic, an extension is a set of sentences in a modal language. We have shown that AE extensions are always stable sets, and so are uniquely determined by their kernel, the set of ordinary formulas of the extension (see Propositions 2.5 and 2.6); further, this kernel is closed under first-order consequence. In default logic, an extension is also a set of first-order sentences closed under FO-consequence. If we choose the first-order language of default logic to be the same as the base langauge $\mathcal{L}_0$ of AE logic, then there is a natural notion of equivalence between the two: we will say that an AE extension is the same as a default extension if the latter is the kernel of the former.

There are now two questions that we wish to answer.

- Is an arbitrary default theory expressible in AE logic?
- Is an arbitrary AE base set expressible as a default theory?

To show that default theories are expressible in AE logic, we must provide a translation from an arbitrary default theory U to an AE base set A, such that U and

A have equivalent extensions. For AE logic, we have three choices for the class of extensions, depending on the groundedness condition — weak, moderate, or strong. The best fit with default logic occurs for strong groundedness, and we define equivalence of base sets accordingly: U and A are equivalent if every extension of U is the same as some strongly-grounded extension of A, and every strongly-grounded extension of A is the same as some extension of U. Obviously, U and A must have the same number of extensions if they are equivalent.

We may want to impose some restrictions on the translation from U to A. The strongest and most natural condition is that it be effectively computable, one-to-one, and context-independent:— every sentence or default rule of U is effectively translated into exactly one sentence of A, independent of any other sentences or defaults in U. Let us call a translation of this sort *local*.[6] The translation that we propose, which follows from our reasoning in section 1 about the relationship between default rules and AE sentences, is a local one. The main result of this paper is that this translation yields an equivalent AE base set.

What about expressing AE base sets as default theories? Here we might expect more difficulty because the structure of default rules is fixed, while the modal operator of AE logic can be arbitrarily embedded in a sentence. However, by reducing a base set A to the normal form developed in section 3.2, we can find a local translation for it into an equivalent default theory U. So if we restrict our attention to the strongly-grounded extensions, AE logic and default logic turn out to be essentially the same, although their formal structure is very different.

5.2 Defaults as self-belief

We now define a local transformation from a default theory $\langle W, D\rangle$ to a set of AE sentences A, such that the default extensions of $\langle W, D\rangle$ are exactly the kernels (the first-order part) of the strongly-grounded AE extensions of A. Thus (as we prove), there is an exact correspondence between default extensions for $\langle W, D\rangle$ and strongly-grounded AE extensions for A.

The transformation is:

$$\frac{\alpha : M\beta_1 \ldots M\beta_n}{\omega} \quad \mapsto \quad (L\alpha \wedge \neg L\neg\beta_1 \wedge \cdots \wedge \neg L\neg\beta_n) \supset \omega \,. \tag{16}$$

As we mentioned in the introduction, this is the natural interpretation of defaults in terms of introspective knowledge. A paraphrase of the AE sentence for agent

[6] Imielinski [6] defines the notion of a *modular* translation: the translation of a default rule in $\langle W, D\rangle$ cannot depend on W, but can depend on the presence of other defaults in D. All local translations are modular.

might be: "If I know that α is true, *and I have no knowledge that any of the* β_i *are false*, then ω must be true." The key phrase has been emphasized; it is in reasoning about what is not known that the nonmonotonic character of AE logic appears. However, the role of the other parts of the sentence ($L\alpha$ and ω) also deserves closer scrutiny; for example, why does w appear as the consequent, and not $L\omega$? From a technical point of view, the transformation (16) is the one that gives a correspondence between default and AE extensions. We will comment more extensively on the intuitions behind the exact form of the transformation, after the basic results are presented.

For most of the rest of this section, we assume that there is at most a single operator $M\beta$ in the antecedent of rules. The proofs can be stated much more simply, and the needed modifications for the general case are obvious. In a default, we allow either α or $M\beta$ to be missing; the corresponding AE sentence just deletes the appropriate conjunct in the antecedent. The conclusion of the default must always be present (defaults with no conclusion are senseless). Let D' be the set of sentences formed by taking the transforms of defaults D; we call the set $\{W, D'\}$ the *AE transform of* $\langle W, D\rangle$.

Now consider a particular default theory $\langle W, D\}$ and an associated extension $E = \Gamma(E)$. E is closed under first-order consequence, and hence is the kernel of a unique stable set. This stable set is closely related to the AE transform of $\langle W, D\rangle$: it is a minimal stable set containing the AE transform. We prove this result as the following proposition.

PROPOSITION 5.1 *Let $\langle W, D\rangle$ be a default theory, with $A = \{W, D'\}$ its AE transform. Suppose E is an extension of the default theory. Then E is the kernel of a minimal stable set containing A and $\neg L\overline{E}$.*

The minimal stable set of the above proposition is also an AE extension of A (recall that a stable set must be (weakly) grounded in A if it is to be an extension of A).

PROPOSITION 5.2 *Same conditions as the previous proposition. The set E is the kernel of a minimal AE extension of A.*

The minimal stable set is also strongly grounded in A, if we consider the obvious normal form for A (that is, the translation of a default rule given by Equation (16) is $\neg L\alpha \vee L\neg\beta_1 \vee \cdots \vee L\neg\beta_n \vee \omega$).

PROPOSITION 5.3 *Same conditions as the previous proposition. The set E is the kernel of a strongly-grounded AE extension of A.*

The converse of this proposition is also true.

PROPOSITION 5.4 *Let A be the AE transform of a default theory $\langle W, D\rangle$. If E is the kernel of a strongly-grounded AE extension of A, then $E = \Gamma(E)$.*

We collect the preceding two propositions into the following theorem, the main result connecting AE and default extensions.

THEOREM 5.5 *Let A be the AE transform of a default theory Δ. A set E is a default extension of Δ if and only if it is the kernel of a strongly-grounded AE extension of A.*

Thus every default theory can be translated into an equivalent AE base set. The extensions of the default theory correspond to the AE belief sets with the strongest groundedness conditions.

5.3 Semantics

By virtue of the translation from default theories to AE logic, we are able to import the semantics of AE logic in analyzing default theories. The semantics of AE sentences is an interpretive semantics, in the sense that a sentence ϕ is true or false in an interpretation $\models_{I,\Gamma}$. The interpretation of modal atoms is given by the modal index Γ, according to Equation (8). The interpretations themselves are straightforward augmentations of standard first-order interpretations. The troublesome characteristics of AE logic, from a semantical point of view, occur in the fixed-point definition of extensions (Definition 2.1), in which only interpretations containing a certain modal index are considered. So, although it is hard to construct and analyze extensions, all of our ordinary intuitions about the meaning of the language $\mathcal{L}$ and its semantics with respect to individual interpretations is still available.

As an example, consider the difference between the two default sentences

$$LBird(Tweety) \wedge \neg L\neg Fly(Tweety) \supset Fly(Tweety) \tag{17}$$

and

$$Bird(Tweety) \wedge \neg L\neg Fly(Tweety) \supset Fly(Tweety)\,. \tag{18}$$

The first sentence states that in any interpretation in which $Bird(Tweety)$ is a belief and $\neg Fly(Tweety)$ is not a belief, $Fly(Tweety)$ will be true. In default logic, the rule which translates to this sentence is $Bird(Tweety) : MFly(Tweety) \,/\, Fly(Tweety)$. The antecedent of the second sentence is less strict: it states only that $Bird(Tweety)$ must be true. The second sentence permits case analysis of a type not sanctioned by

the first. For example, suppose it is known that either Tweety is a bird, or Tweety is housebroken ($Houseb(Tweety)$). In every interpretation in which $\neg Fly(Tweety)$ is not a belief and the second default sentence is true, $Houseb(Tweety) \vee Fly(Tweety)$ is true. On the other hand, nothing can be concluded by assuming the first default sentence is true, because $Bird(Tweety)$ itself may not be a belief.[7]

The distinction between (17) and (18) makes clear the reason why $L\alpha$ must be used to translate default rules. We can also answer a question posed earlier: why does ω appear in the antecedent of the AE translation, instead of $L\omega$? The reason is that, from the agent's point of view, the conclusion ω is the simple belief that ω is true of the world. Using $L\omega$ would mean that the agent concludes she has a self-belief $L\omega$, and she would have to reason further from that to the simple belief in ω. As we have pointed out in the discussions on moderate and strong groundedness, such reasoning can lead to ungrounded justification of beliefs, and is to be avoided.

Another instance of the utility of interpretive semantics is in the concepts of equivalence and substitution. Two formulas ϕ and ϕ' of $\mathcal{L}$ are equivalent if they have the same truthvalue in all models. Because the definition of AE extensions is framed in terms of the interpretive semantics, ϕ' can be substituted wherever ϕ occurs in a base set A, without changing the AE extensions of A. We used this fact extensively in arriving at the normal form for AE sentences in section 3.2.

For a final example of the use of equivalence, we turn to the literature of inheritance networks. These networks are meant to express class/subclass relationships, and the inheritance of properties by default from a class to its subclasses. Touretzky (in [17, p. 34]) has offered a translation from his particular type of inheritance networks into AE logic. To formally express the intent of the statement "P implies that typically Q is unknown," he uses:

$$P \wedge \neg L(L\neg Q \vee LQ) \supset \neg L\neg Q \wedge \neg LQ \tag{19}$$

However, from Example 3.1 we know that this is equivalent to a tautology, and thus not a very satisfying translation.

[7] Etherington [3, p. 34] cites this as evidence that the second sentence seems more in accord with our intuitions about the way defaults should work. However, there are some objectionable consequences of using (18) as the formal representation of a default. For example, by simple propositional manipulations, it can also be viewed as a default stating that nonflying things are not birds (Matt Ginsberg originally pointed this out to the author). As can be seen from this example, the question of the appropriateness of the formal system for representing our intuitions about defaults is a complicated one, and is not a direct concern of this paper. However, a clear understanding of the formal consequences of the representation can facilitate this discussion.

5.4 Expressiveness

The question of expressiveness can be phrased as follows: Is it the case that default sentences of the type (18), or perhaps other AE sentences involving complicated constructions such as embedded L operators, have no counterpart in default theories? On the face of it this would seem a plausible conjecture, since the L operator is part of the language, while default rules are not. However, it turns out that AE logic is no more expressive than default logic: there is an effective transformation of any base set of AE sentences into a default theory, such that the default extensions are exactly the kernels of the strongly-grounded AE extensions. To show this, we rely on the fact (see Proposition 3.9) that every set of sentences of $\mathcal{L}$ has an equivalent normal form in which every sentence looks like:

$$\neg L\alpha \vee L\beta_1 \vee \cdots \vee L\beta_n \vee \omega \ , \tag{20}$$

where all of α, β_i, and ω are ordinary sentences. Any of the modal atoms may be missing, but ω is always present.

Given any set of $\mathcal{L}$-sentences A in normal form, it is possible to effectively construct a corresponding default theory (W, D), in the following way. Any ω that appears without other disjuncts is put into W. All other sentences are transformed into defaults, using the converse of by Equation 16:

$$\neg L\alpha \vee L\beta_1 \vee \cdots \vee L\beta_n \vee \omega \quad \mapsto \quad \frac{\alpha : M\neg\beta_1 \ldots M\neg\beta_n}{\omega} \ . \tag{21}$$

There is one slight asymmetry here, however. A normal form AE sentence could be missing $\neg L\alpha$, or all of the $L\beta_i$'s. There is no provision in default rules for omitting any part of the premisses. So we define an *extended normal form* by first putting a set of AE sentences into normal form, and then using the following two translations to add disjuncts where necessary:

$$\begin{array}{lcl} L\beta_1 \vee \cdots \vee L\beta_n \vee \omega & \mapsto & \neg L\top \vee L\beta_1 \vee \cdots \vee L\beta_n \vee \omega \\ \neg L\alpha \vee \omega & \mapsto & \neg L\alpha \vee L\bot \vee \omega \end{array} \tag{22}$$

It is not hard to show that the *consistent* strongly grounded extensions of a set A in normal form are the same as those of its extended normal form A'. However, A may have an inconsistent extension which is not realized by A': take $A = \{P, \neg LP \vee \neg P\}$, for example, which has the inconsistent stable set as its one strongly grounded extension. The extended normal form $A' = \{P, \neg LP \vee L\bot \vee P\}$ has no strongly-grounded extensions.

For any set of sentences A in extended normal form, we define a corresponding default theory (W, D) using Equation (21). It is easy to see that A is the AE

transform of $\langle W, D\rangle$; by Theorem 5.5, these two have essentially the same extensions. More precisely, we have proven the following theorem:

THEOREM 5.6 *For any set of sentences A of $\mathcal{L}$ in extended normal form, there is an effectively constructable default theory $\langle W, D\rangle$ such that E is a default extension of $\langle W, D\rangle$ if and only if it is the kernel of a strongly-grounded extension of A.*

So, suprisingly, default theories have the same expressiveness as AE logic over the modal language $\mathcal{L}$. However, some caveats should be noted. The extensions are the same only if we restrict ourselves to strongly-grounded ones on the AE side. If we wish to use, say, moderate groundedness as our condition on ideal belief sets, then the translation into default logic is lacking in that some of the AE extensions may not have corresponding default extensions.

A second caveat is that if we extend $\mathcal{L}$ by allowing quantifying-in (i.e., expressions such as $\exists x.L\phi(x)$), in all likelihood Theorem 5.6 will no longer hold. There are a number of reasons to think this; perhaps the most compelling is Levesque's observation [7] that in the presence of quantifying-in, there are sentences of modal depth greater than 1 with no equivalents in $\mathcal{L}_1$.

Given that AE logic can be embedded in default logic, we can translate various types of defaults that have a natural expression in AE logic. The default rules corresponding to the two types of defaults (17) and (18) are

$$\frac{Bird(Tweety) : MFly(Tweety)}{Fly(Tweety)} \tag{23}$$

and

$$\frac{MFly(Tweety)}{Bird(Tweety) \supset Fly(Tweety)} . \tag{24}$$

Note that the second type of default (Equation 18), when translated into the default rule (24), causes the atom $Bird(Tweety)$ to appear not in the antecedent of the default, but in the consequent.

Conditional defaults are expressed in $\mathcal{L}$ by sentences of the form:

$$C \supset (L\alpha \wedge \neg L\neg\beta \supset \omega) \tag{25}$$

By simple propositional manipulations, this is equivalent to:

$$\neg L\alpha \vee L\neg\beta \vee (C \supset \omega) , \tag{26}$$

which translates into the default rule

$$\frac{\alpha : M\beta}{C \supset \omega} . \tag{27}$$

The condition is expressed in the consequent of the default rule.

A default whose conclusion is a default can be expressed by:

$$L\alpha \wedge \neg L\neg\beta \supset (L\alpha' \wedge \neg L\neg\beta' \supset \omega') \tag{28}$$

Again, by propositional manipulations this can be put into the form

$$\neg L\alpha \vee \neg L\alpha' \vee L\neg\beta \vee L\neg\beta' \vee \omega' . \tag{29}$$

The first two disjuncts can be combined into the $K45$-equivalent literal $\neg L(\alpha \wedge \alpha')$. The resulting default rule translation is

$$\frac{\alpha \wedge \alpha' : M\beta,\ M\beta'}{\omega} . \tag{30}$$

6 Conclusion

Given the current proliferation of nonmonotonic formalisms, it seems wise to establish comparisons among them, especially regarding expressiveness. The results presented here show that there is an exact correspondence between AE logic over $\mathcal{L}$ and default theories. There is a general, effective translation between the two that is *local*: each sentence (or default rule) can be translated in isolation from the others. The translation preserves theoremhood, in that the default extensions are the first-order part of the strongly-grounded AE extensions.

The relationship between default logic and various forms of circumscription [10] has been investigated by a number of researchers. Etherington [3] collects a number of comparability results, and notes that there is a fundamental difference between the minimal-model semantics of circumscription and the fixed-point semantics of default theories. In particular, he cites the following points:

1. Default logic is *brave* in the sense that in the presence of competing defaults, individual extensions will satisfy a maximally consistent set of defaults; in contrast, a natural corresponding circumscriptive theory for defaults would be *cautious*, inferring only what the competing extensions have in common.

2. Default logic can make nonmonotonic inferences about equality, and circumscription cannot.

3. Circumscriptive statements can apply to all individuals, whereas default rules are restricted to those individuals with names in the language.

4. In default logic, there seems to be no way to capture the distinction between *fixed* and *variable* predicates in circumscription.

These results suggest that circumscription and default logic are incomparable, in the sense that there is no local translation of one into the other that preserves theoremhood.

What about the relationship between AE logic and circumscription? Because AE logic is "brave," and can make nonmonotonic inferences about equality (if we allow an equality predicate in $\mathcal{L}$), there is no local translation of AE theories into circumscription. And as long as the language of AE logic is $\mathcal{L}$, it will have the same expressiveness as default logic, and the last two items seem to preclude any general translation of circumscription into AE logic. By allowing quantifying-in in the language, it is possible to construct AE statements applying to all individuals, and so the third item may not present a problem. The fourth item remains, however, and creates pessimism about the existence of a translation from circumscription with fixed predicates.

7 Acknowledgements

This paper is a rewritten and expanded version of a preliminary report that appeared in IJCAI87 in Milan. I would like to thank Michael Gelfond, Hector Levesque, W. Marek, Karen Myers, Donald Perlis, Halina Przymusinska, Ray Reiter, and M. Truszczynski for their careful reading and comments on the paper. In particular, Michael Gelfond and Halina Przymusinski found a counterexample to the original "proof" of the equivalence of minimal AE extensions and default theories, which led me to formulate the concept of strongly-grounded extensions, and W. Marek and M. Truszczynski pointed out a needed modification to the definition of minimal AE extension.

References

[1] Boolos, G. *The Unprovability of Consistency.* Cambridge University Press, Cambridge, England, 1979.

[2] Chellas, B. F. *Modal Logic: An Introduction.* Cambridge University Press, 1980.

[3] Etherington, D. W. *Reasoning with Incomplete Information: Investigations of Non-Monotonic Reasoning.* PhD thesis, University of British Columbia, Vancouver, British Columbia, 1986.

[4] Halpern, J. Y. and Moses, Y. O. A guide to the modal logics of knowledge and belief. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 50–61, Los Angeles, 1985.

[5] Hintikka, J. Impossible possible worlds vindicated. *Journal of Philosophical Logic*, 4:475–484, 1975.

[6] Imielinski, T. Results on translating defaults to circumscription. *Artificial Intelligence*, 32(1):131–146, April 1987.

[7] Levesque, H. J. *A Formal Treatment of Incomplete Knowledge Bases.* Technical Report 614, Fairchild Artificial Intelligence Laboratory, Palo Alto, California, 1982.

[8] Łukaszewicz, W. Two results on default logic. In *Proceedings of the American Association of Artificial Intelligence*, pages 459–461, University of California at Los Angeles, 1985.

[9] Marek, W. Stable theories in autoepistemic logic. 1986. Unpublished Note, Department of Computer Science, University of Kentucky.

[10] McCarthy, J. Circumscription — a form of nonmonotonic reasoning. *Artificial Intelligence*, 13(1–2), 1980.

[11] Moore, R. C. *Possible-world Semantics for Autoepistemic Logic.* Technical Note 337, SRI Artificial Intelligence Center, Menlo Park, California, 1984.

[12] Moore, R. C. *Reasoning from Incomplete Knowledge in a Procedural Deduction System.* Technical Report AI–TR–347, MIT Artificial Intelligence Laboratory, 1975.

[13] Moore, R. C. Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, 25(1), 1985.

[14] Reiter, R. A logic for default reasoning. *Artificial Intelligence*, 13(1–2), 1980.

[15] Robinson, J. A. *Logic: Form and Function.* Elsevier North Holland, New York, 1979.

[16] Stalnaker, R. C. A note on nonmonotonic modal logic. 1980. Department of Philosophy, Cornell University.

[17] Touretzky, D. S. *The Mathematics of Inheritance Systems.* Morgan Kaufmann Publishers, Inc., Los Altos, California, 1986.

Appendix: Propositions and Proofs

This appendix contains all theorems, propositions, lemmas, and their proofs. Some informal proofs have already been given in the main body of this paper, and are not repeated here.

PROPOSITION 2.1 *If A is a set of ordinary sentences, it has exactly one AE extension T. T_0 is the first-order closure of A.*

Proof. We define the sets $S(n)$ in the following iterative fashion:

$$\begin{aligned} S(0) &= \{\phi \in \mathcal{L}_0 \mid A \models \phi\} \\ S(n) &= \{\phi \in \mathcal{L}_n \mid A \models_{S(n-1)} \phi\} \end{aligned}$$

Let T_n be the set of sentences of T from $\mathcal{L}_n$, and let S be the infinite union of all $S(i)$. We will show that if T is an AE extension of A, $T_n = S(n)$; there is thus at most one AE extension T, with $T_0 = S(0)$, the first-order closure of A. We prove existence by showing that S is always an AE extension of A.

Let T be an AE extension of A. Obviously, $T_0 = S(0)$. Assume that $T_{n-1} = S(n-1)$. We have:

$$\begin{aligned} T_n &= \{\phi \in \mathcal{L}_n \mid A \models_T \phi\} \\ &= \{\phi \in \mathcal{L}_n \mid A \models_{T_{n-1}} \phi\} \\ &= \{\phi \in \mathcal{L}_n \mid A \models_{S(n-1)} \phi\} \end{aligned}$$

The second equality holds because the truthvalue of any sentence in $\mathcal{L}_n$ depends only on the subset of the modal index in $\mathcal{L}_{n-1}$.

Let us define S' by:

$$S' = \{\phi \mid A \models_S \phi\} .$$

To show that S is an AE extension of A, we will show by induction that $S = S'$. Let us assume for the moment that $S_n = S(n)$, that is, S restricted to $\mathcal{L}_n$ is just the set $S(n)$. For the base step, it is obvious that $S'_0 = S(0) = S_0$. Now assume that $S'_{n-1} = S_{n-1}$. Then we have:

$$\begin{aligned} S'_n &= \{\phi \in \mathcal{L}_n \mid A \models_S \phi\} \\ &= \{\phi \in \mathcal{L}_n \mid A \models_{S_{n-1}} \phi\} \\ &= \{\phi \in \mathcal{L}_n \mid A \models_{S(n-1)} \phi\} \\ &= S(n) \\ &= S_n \end{aligned}$$

Now we show that $S_n = S(n)$. As preliminary facts, observe that:

$$\begin{aligned} S(k) &= \{\phi \in \mathcal{L}_k \mid A \models_{S(k-1)} \phi\} \\ &= \{\phi \in \mathcal{L}_k \mid A \models_{S(k)} \phi\} \\ &= S(k+1)_k \ , \end{aligned}$$

and hence, by simple induction, for any k and $j > 0$, $S(k) = S(k+j)_k$.

We already know that $S(n) \subseteq S_n$. Assume that S_n is larger than $S(n)$, so there is a sentence ϕ such that $\phi \in S_n$ and $\phi \notin S(n)$. Let the modal depth of ϕ be k, where $k \leq n$. For all $j > 0$, $S(k+j)_k = S(k)$, and so ϕ must be in $S(n)$, a contradiction. Therefore S_n cannot be larger than $S(n)$, and must be equal to it.

PROPOSITION 2.2 *A set T is an AE extension of A if and only if it satisfies the equation*

$$T = \{\phi \mid A \cup LT \cup \neg L\overline{T} \models \phi\} \ .$$

PROPOSITION 2.3 (MOORE)

Every AE extension of A is a stable set containing A.

PROPOSITION 2.4 *Every stable set Γ is an AE extension of Γ_0.*

Proof. For any set Γ, it must be the case that

$$\Gamma = \{\phi \mid \Gamma \models \phi\} \ .$$

If Γ is stable, then both $L\Gamma \subset \Gamma$ and $\neg L\overline{\Gamma} \subset \Gamma$, so we have

$$\Gamma = \{\phi \mid \Gamma \cup L\Gamma \cup \neg L\overline{\Gamma} \models \phi\} \ .$$

By the argument preceding proposition 2.2, we know from this that

$$\Gamma = \{\phi \mid \Gamma \models_\Gamma \phi\} \ .$$

Finally, only the ordinary sentences of Γ are important in determining whether $\models_{I,\Gamma} \Gamma$ is true, and so

$$\Gamma = \{\phi \mid \Gamma_0 \models_\Gamma \phi\} \ .$$

PROPOSITION 2.5 (MOORE)

If two stable sets agree on ordinary formulas, they are equal.

PROPOSITION 2.6 *Let W be a set of ordinary formulas closed under first-order consequence. There is a unique stable set Γ such that $\Gamma_0 = W$. W is called the* kernel *of the stable set.*

Proof. By Proposition 2.1, there is a unique AE extension T of W. T is a stable set (Proposition 2.3), and by the discussion preceding Proposition 2.1, T_0 is the set of ordinary sentences logically implied by W. Since W is FO-closed, $T_0 = W$.

PROPOSITION 2.7 *A set T is an AE extension of A if and only if it satisfies the equation*

$$T = \{\phi \mid A \cup LT_0 \cup \neg L\overline{T}_0 \models_{SS} \phi\} .$$

Proof. If T is an AE extension, then by Proposition 2.3 it must be stable, and from the above discussion the fixed-point equation reduces to

$$T = \{\phi \mid A \models_T \phi\} ,$$

which is just the definition of an AE extension of A.

On the other hand, suppose T satisfies the fixed-point equation. T_0 is a set of ordinary formulas closed under first-order consequence; hence by Proposition 2.6 they are exactly the ordinary formulas of a unique stable set; so LT_0 and $\neg L\overline{T}_0$ determine a unique stable set, which we call S. The fixed-point equation reduces to

$$T = \{\phi \mid A \models_S \phi\} .$$

T will be an AE extension of A if we can show that $T = S$; note that we already know that the kernels of T and S are equal.

We define the set A' as the the set of sentences in A, where $L\phi$ is replaced everywhere by $\top$ if $\phi \in S$, and by $\bot$ if $\phi \notin S$. By the truth-recursion rules for L valuations, it must be the case that $\models_{I,S} A$ if and only if $\models_{I,S} A'$. The sentences of A' are all ordinary, and their FO closure is just T_0. Hence $\models_{I,S} A$ if and only if $\models_{I,S} T_0$. Since $T_0 = S_0$, we have

$$T = \{\phi \mid S_0 \models_S \phi\} .$$

By Proposition 2.4, the right-hand side defines the stable set S, and so $T = S$.

PROPOSITION 2.8 *A set of sentences is moderately grounded in A if and only if it is a minimal AE extension of A.*

Proof. Assume that T is a minimal AE extension of A. By definition, it obeys the equation

$$T = \{\phi \mid A \models_T \phi\} .$$

Because T is minimal for A, there is no other stable set containing A which does not contain some element of $\overline{T}_0$. Hence $LA \cup \neg L\overline{T}_0$ determine the unique stable set T, and the above equation can be rewritten as

$$T = \{\phi \mid A \cup LA \cup \neg L\overline{T}_0 \models_{SS} \phi\} .$$

On the other hand, assume that T is moderately grounded in A. First we show that if $\phi \in T$, where ϕ is ordinary, then $L\phi \in T$. Consider any L valuation $\langle I, \Gamma \rangle$ that satisfies $A \cup LA \cup \neg L\overline{T}_0$. By the properties of stable sets, $A \cup LA \cup \neg L\overline{T}_0 \subset \Gamma$. By Propositions 2.4 and 2.7, we have

$$\begin{aligned} \Gamma &= \{\phi \mid \Gamma_0 \models_\Gamma \phi\} \\ &= \{\phi \mid \Gamma_0 \cup L\Gamma_0 \cup \neg L\overline{\Gamma}_0 \models_{SS} \phi\} , \end{aligned}$$

so we must have $\Gamma_0 \cup L\Gamma_0 \cup \neg L\overline{\Gamma}_0 \models_{SS} A \cup LA \cup \neg L\overline{T}_0$; and since we know that $A \cup LA \cup \neg L\overline{T}_0 \models_{SS} \phi$, it must be that $\phi \in \Gamma$. Since Γ was arbitrarily chosen, $A \cup LA \cup \neg L\overline{T}_0 \models_{SS} L\phi$.

Now the proof proceeds along the lines of Proposition 2.7. We have

$$T = \{\phi \mid A \cup LA \cup LT_0 \cup \neg L\overline{T}_0 \models_{SS} \phi\} ,$$

and because T_0 is closed under first-order consequence, by Proposition 2.6 it determines a unique stable set T' with $T'_0 = T_0$. We rewrite the above equation as

$$T = \{\phi \mid A \cup LA \models_{T'} \phi\} .$$

The assumption LA is irrelevant, because $A \subset T'$; hence, by the reasoning in the proof of Proposition 2.7,

$$T = \{\phi \mid T'_0 \models_{T'} \phi\} ,$$

which just defines the stable set T', and $T = T'$. T is thus an AE extension of A, since

$$T = \{\phi \mid A \models_T \phi\} .$$

T must also be minimal for A; if it weren't, then there would be another stable set containing A (and hence LA) and not containing any of $\overline{T}_0$; hence $A \cup LA \cup \neg L\overline{T}_0$ would not determine a unique stable set, which we have shown to be the case.

PROPOSITION 3.1 (MOORE)

A set T is an AE extension of A if and only if it satisfies the equation

$$T = \{\phi \mid A \cup LT \cup \neg L\overline{T} \vdash \phi\} .$$

PROPOSITION 3.2 *A set is stable if and only if it is an S5 set.*

PROPOSITION 3.3 *For any $\phi \in \mathcal{L}$, $\models_{S5^+} \phi$ if and only if $\models_{S5} \phi$.*

Proof. This proposition is true if every $S5^+$ valuation has an equivalent L valuation whose modal index is a stable set, and *vice versa*. Let $\langle w, W\rangle$ be an $S5^+$ valuation, and Γ the $S5$ set defined by W. By the truth-recursion definition (11), $\models_{S5^+} L\phi$ if and only if $\phi \in \Gamma$. Hence $\langle w, W\rangle$ is equivalent to the L valuation $\langle w, \Gamma\rangle$. By Proposition 3.2, Γ is a stable set.

On the other hand, let $\langle I, \Gamma\rangle$ be an L valuation with Γ stable. Γ is also an $S5$ set; let W be a set of possible worlds such that Γ_0 are exactly the ordinary sentences true at every world in the set. Then $\models_{\langle I,W\rangle} L\phi$ if and only if $\models_{\langle I,\Gamma\rangle} L\phi$, and these are equivalent valuations.

LEMMA 3.4 *Let R be an equivalence relation on W, and let the successors of w be the subset $W' \subseteq W$. Then the Kripke valuation $\langle w, W, R\rangle$ is equivalent to the $S5^+$ valuation $\langle w, W'\rangle$.*

Proof. The lemma will be true if the truth-recursion schemes (12) and (11) are identical. The first parts obviously are. For the second part, note that the elements w' such that wRw' are exactly the set W'. Hence the second clause of (12) can be rewritten to be identical with that of (11).

PROPOSITION 3.5 *A valuation is an $S5^+$ valuation if and only if it is a TE valuation.*

Proof. Let $\langle w, W, R\rangle$ be a TE valuation. Let W' be all successors of w, that is, the set of w' such that wRw'. Because R is euclidean, it is an equivalence relation on W' (any two elements a and b of W' have wRa and wRb, and by the euclidean condition aRb is true). By the lemma above, $\langle w, W, R\rangle$ is equivalent to the $S5^+$ valuation $\langle w, W'\rangle$.

Let $\langle w, W\rangle$ be an $S5^+$ valuation. Let R be a relation that is an equivalence on W, and also for any element $a \in W$, wRa. By the

lemma above, $\langle w, W\rangle$ is equivalent to the Kripke valuation $\langle w, W, R\rangle$. We must now show that R is transitive and euclidean.

Let aRb and bRc. If $a \in W$, then aRc because R is an equivalence on W. If $a \notin W$, it must be w, so that $b \in W$ and $c \in W$, and bRc by the equivalence of R on W.

Let aRb and aRc. If $a \in W$, then $b \in W$ and $c \in W$ by the equivalence property of R, and hence bRc. If $a \notin W$, it must be w, and again $b \in W$ and $c \in W$.

PROPOSITION 3.6 *A set T is an AE extension of A if and only if it satisfies the equation*

$$T = \{\phi \mid A \cup LT_0 \cup \neg L\overline{T}_0 \vdash_{K45} \phi\} .$$

It is a minimal (moderately-grounded) extension of A if and only if it satisfies the equation

$$T = \{\phi \mid A \cup LA \cup \neg L\overline{T}_0 \vdash_{K45} \phi\} .$$

PROPOSITION 3.7 *Every sentence of $\mathcal{L}_1$ is equivalent to a sentence of the form*

$$\begin{array}{ll} (L_1 \vee \omega_1) & \wedge \\ (L_2 \vee \omega_2) & \wedge \\ \cdots & \wedge \\ (L_n \vee \omega_n) & \end{array}$$

where each L_i is a disjunction of modal literals on ordinary sentences, and each ω_i is ordinary.

Proof. Using only first-order valid operations, a sentence of $\mathcal{L}_1$ is put into prenex normal form:

$$Q_1 x_1 \ldots Q_n x_n \, M \, , \tag{31}$$

where Q_i is either $\forall$ or $\exists$, and M is a propositional matrix containing modal and ordinary atoms. Note that the modal atoms themselves may contain embedded quantifiers, e.g., $L\exists x P(x)$.

We will use the following equivalences:

$$\begin{aligned} Qx([\neg]L\phi \vee \psi) &\equiv [\neg]L\phi \vee Qx\psi \\ Qx([\neg]L\phi \wedge \psi) &\equiv [\neg]L\phi \wedge Qx\psi \\ \exists x(\phi \vee \psi) &\equiv \exists x\phi \vee \exists x\psi \\ \forall x(\phi \wedge \psi) &\equiv \forall x\phi \wedge \forall x\psi \end{aligned}$$

where $[\neg]L\phi$ is either $L\phi$ or $\neg L\phi$. The first two are consequences of the fact that $L\phi$ contains no free variables. The second two are first-order consequences (see Robinson [15]).

Returning to the prenex form (31): suppose Q_n is an existential quantifier. Then put M into disjunctive normal form, and using the equivalence for $\exists x$, distribute $\exists x$ onto each of the disjuncts. The matrix now looks like

$$\begin{array}{ll} \exists x(L_1 \wedge \omega_1) & \vee \\ \exists x(L_2 \wedge \omega_2) & \vee \\ \cdots & \vee \end{array} \tag{32}$$

where L_i are conjunctions of modal literals and ω_i are ordinary. By using the first two equivalences, the quantifiers can be pushed through the modal literals (modal atoms or their negations), giving:

$$\begin{array}{ll} L_1 \wedge \exists x\, \omega_1 & \vee \\ L_2 \wedge \exists x\, \omega_2 & \vee \\ \cdots & \vee \end{array} \tag{33}$$

If Q_n is a universal quantifier, then the same operations can take place, except that the matrix is put into conjunctive normal form.

The same process can be repeated for Q_{n-1}, where we now consider expressions of the form $Q_n x_n \omega_i$ to be atomic for the purposes of putting M into conjunctive or disjunctive normal form. In this way all quantifiers can be pushed around the modal atoms, and the end result is a sentence of the form

$$\begin{array}{ll} (L_1 \vee \omega_1) & \wedge \\ (L_2 \vee \omega_2) & \wedge \\ \cdots & \wedge \\ (L_n \vee \omega_n) & \end{array} \tag{34}$$

where each L_i is a disjunct of modal literals and each ω_i is ordinary.

PROPOSITION 3.8 *Every modal atom $L\phi$, where ϕ is from $\mathcal{L}_1$, is equivalent to a sentence of $\mathcal{L}_1$.*

Proof. We will use the following equivalences:

$$\begin{aligned} L(\phi \wedge \psi) &\equiv L\phi \wedge L\psi \\ LL\phi &\equiv L\phi \\ L\neg L\phi &\equiv \neg L\phi \vee L\bot \\ L(L\phi \vee \psi) &\equiv L\phi \vee L\psi \\ L(\neg L\phi \vee \psi) &\equiv \neg L\phi \vee L\bot \vee L\psi \end{aligned}$$

All of these are theorems of *K45*.

By Proposition 3.7 and the first equivalence above, a modal atom $L\phi$, where ϕ is in $\mathcal{L}_1$, can be put into the equivalent form

$$\begin{array}{ll} L(L_1 \vee \omega_1) & \wedge \\ L(L_2 \vee \omega_2) & \wedge \\ \cdots & \wedge \\ L(L_n \vee \omega_n) & \end{array}$$

where each L_i is a disjunction of modal literals on ordinary sentences, and each ω_i is ordinary. By applying the equivalences repeatedly, this can be reduced to a sentence in $\mathcal{L}_1$.

PROPOSITION 3.9 *Every set A of $\mathcal{L}$-sentences has a K45-equivalent set in which each sentence is of the form*

$$\neg L\alpha \vee L\beta_1 \vee \cdots \vee L\beta_n \vee \omega \ , \tag{35}$$

with α, β_i, and ω all being ordinary sentences. Any of the disjuncts, except for ω, may be absent.

PROPOSITION 3.10 *If T is a strongly grounded extension of A, it is moderately grounded in A.*

Proof. Since T contains A and LA, we can add these as premises to the fixed-point equation for strongly-grounded T:

$$T = \{\phi \mid A \cup LA \cup A' \cup LA' \cup \neg L\overline{T}_0 \models_{SS} \phi\} \ .$$

Since A subsumes A' and LA subsumes LA', this reduces to the equation for a moderately-grounded extension.

PROPOSITION 5.1 *Let $\langle W, D\rangle$ be a default theory, with $A = \{W, D'\}$ its AE transform. Suppose E is an extension of the default theory. Then E is the kernel of a minimal stable set containing A and $\neg L\overline{E}$.*

Proof. The stable set S whose kernel is E contains W. Consider the AE transform $L\alpha \wedge \neg L\neg\beta \supset \omega$ of an arbitrary member d of D, and suppose it is not in S. Then ω is not in E, α is in E, and $\neg\beta$ is not in E. But this means that d is not satisfied by $\Gamma(E)$, a contradiction. Hence S contains all of D', the AE transform of D. Because its kernel is E, it also contains $\neg L\overline{E}$.

To show that S is minimal, assume that there is some other set $V \subset E$ that is the kernel of a stable set containing A and $\neg L\overline{E}$. Consider an arbitrary member $L\alpha \wedge \neg L\neg\beta \supset \omega$ of D'; because this is a member of the stable set, either α is not V, $\neg\beta$ is in V (and hence E), or ω is in V. If this is the case, then all defaults D of the default theory are satisfied in V, and so $\Gamma(E) \subseteq V$ by the minimality of Γ. But this contradicts $\Gamma(E) = E$.

PROPOSITION 5.2 *Let $\langle W, D\rangle$ be a default theory, with $A = \{W, D'\}$ its AE transform. Suppose E is an extension of the default theory. Then E is the kernel of a minimal AE extension of A.*

Proof. By Proposition 5.1, E is the kernel of a minimal stable set S containing A and $\neg L\overline{E}$. Now consider the set

$$\{\phi | A \cup LA \cup \neg L\overline{E} \models_{SS} \phi\} .$$

We want to show that this set is equal to S. By the preceding proposition, S is a minimal stable set containing A and $\neg L\overline{E}$. In fact, it is the only stable set with this property, because the kernel of any other stable set containing them must be a subset of E, and hence S would not be minimal. Thus LA and $\neg L\overline{E}$ pick out the unique stable set S. We can rewrite the above set as

$$\{\phi | A \models_S \phi\} .$$

Let us call this set T. The modal index, S, is fixed, so we can rewrite it as

$$\{\phi | T_0 \models_S \phi\} .$$

If we can show that $T_0 = E$, then by Proposition 2.4 it must be the case that $T = S$, and S is an AE extension of A. Since it is a minimal stable set containing A, it is also a minimal AE extension, by Definition 2.3.

But T_0 must be the same as E, since it is just the first-order closure of W and the consequents ω of all AE transforms whose antecedents are satisfied by the modal index S. This is exactly the same as $\Gamma(E)$, which by hypothesis is equal to E.

PROPOSITION 5.3 *Same conditions as the previous proposition. E is the kernel of a strongly-grounded AE extension of A.*

Proof. Consider the subset of defaults $D_1 \subseteq D$ whose conclusions are in E. E is still a default extension of $\langle W, D_1 \rangle$. Therefore, by Proposition 5.2, the AE transform $A_1 = \langle W, D_1' \rangle$ has a minimal extension T' whose kernel is E. Since AE extensions are uniquely determined by their kernels, $T' = T$. Thus we have:

$$T = \{\phi \mid A_1 \cup LA_1 \cup \neg L\overline{T}_0 \models_{SS} \phi\} .$$

Since A_1 are all of A whose ordinary part is contained in T, T must be strongly grounded.

PROPOSITION 5.4 *Let A be the AE transform of a default theory $\langle W, D \rangle$. If E is the kernel of a strongly-grounded AE extension of A, then $E = \Gamma(E)$.*

Proof. By Proposition 2.8, a strongly-grounded extension T of A obeys the equation

$$T = \{\phi \mid A_1 \cup LA_1 \cup \neg L\overline{T}_0 \models_{SS} \phi\} ,$$

where A_1 are all of A whose ordinary part is contained in T.

We now show that $\Gamma(T_0)$ cannot be a proper subset of T_0. Assume that this is so. By definition, $\Gamma(T_0)$ satisfies all defaults D. Let $D_1 \subseteq D$ be those defaults whose transform is A_1. Consider the transform $L\alpha \wedge \neg L\neg\beta \supset \omega$ of any member of D_1. Because the default is satisfied, either ω is in $\Gamma(T_0)$ or α is not in $\Gamma(T_0)$ (we already know that $\neg\beta$ is not in T_0). Hence every member of D_1' (and $A_1 = D_1' \cup W$) is in the stable set S whose kernel is $\Gamma(T_0)$. But by Proposition 3.10, T is moderately grounded and hence (by Proposition 2.8) a minimal stable set containing A_1, a contradiction. Thus $\Gamma(T_0)$ is equal to or a superset of T_0.

T_0 is closed under first-order consequence and contains W. It also satisfies all of the defaults D. Consider a default $\alpha : M\beta/\omega$; because its AE transform is in T, and so either ω is in T_0, α is not in T_0, or $\neg\beta$ is in T_0. These are precisely the conditions for the satisfaction of the default.

Since $\Gamma(T_0)$ is the least first-order-closed set containing W and satisfying D, and since it must at least contain T_0, it is equal to T_0. Thus T_0 is a fixed point of Γ and a default extension of $\langle W, D \rangle$.

THEOREM 5.5 *Let A be the AE transform of a default theory Δ. A set E is a default extension of Δ if and only if it is the kernel of a strongly-grounded AE extension of A.*

THEOREM 5.6 *For any set of sentences A of $\mathcal{L}$ in extended normal form, there is an effectively constructable default theory $\langle W, D\rangle$ such that E is a default extension of $\langle W, D\rangle$ if and only if it is the kernel of a strongly-grounded extension of A.*